Introdução aos testes não-paramétricos de tipo ANOVA

Teste de Kruskal-Wallis

[Kruskal, W.H., Wallis, W.A.: Use of ranks in one-criterion variance analysis. J. Am. Stat. Assoc. **47**, 583–621 and errata, ibid. **48**, 907–911 (1952).]

Variante não paramétrica de uma ANOVA a 1 factor (delineamento totalmente casualizado), com k níveis

- Os valores observados são substituídos pelas suas ordens sobre todas as n observações (as ordens serão números naturais $1, 2, \ldots, n$; o valor observado mais baixo é substituído por 1, o maior por n). A solução usual para o teste de Kruskal-Wallis com valores repetidos é atribuir a cada grupo de observações repetidas uma ordem comum dada pela média das ordens que caberiam ao conjunto dessas observações.

- **Para o caso $k = 2$, o teste não paramétrico de Wilcoxon-Mann-Whitney compara hipóteses análogas às do teste de Kruskal-Wallis.**

Teste de Kruskal-Wallis

Hipóteses

Seja $F_i\,(y)$ a função distribuição cumulativa das observações correspondentes ao nível i do factor.

- Hipótese nula:

  H_0: $F_i\,(y) = F(y)\ \ (\forall y \in \mathbb{R}; \forall i = 1, \ldots, k)$

  (Os k níveis do factor possuem idêntica distribuição)

- Hipótese alternativa:

  H_1: $\exists i, j\ :\ F_i\,(y) < F_j\,(y) \qquad (\forall y \in \mathbb{R})$

  (Pelo uma das distribuições dá valores mais elevados do que os das outras)

O teste diz-se não paramétrico porque não é importante conhecer a natureza exacta das distribuições $F_i\,(y)$

Teste de Kruskal-Wallis

Admitamos que não há valores repetidos: os r_{ij} tomam os valores naturais de 1 a $n = \sum_{i=1}^{k} n_i$

Estatística do Teste de Kruskal-Wallis

$$H = \frac{12}{n(n+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(n+1)$$

- $k,$ número de níveis do factor
- $n = \sum_{i=1}^{k} n_i$
- n_i , número de observações no nível i do factor
- $R_i = \sum_{j=1}^{n_i} r_{ij}$, a soma das ordens das n_i observações do nível i

Teste de Kruskal-Wallis

Distribuição assintótica da estatística H

É possível determinar a distribuição exacta de H, admitindo que qualquer possível ordenação das n ordens é igualmente provável. No entanto, a distribuição exacta da estatística não é fácil de calcular. Conover (1980) apresenta um extrato da tabela da distribuição exacta da estatística de Kruskal-Wallis quando k=3.

Mas, para $n_i s$ grandes, e mesmo para $n_i s$ modestos ($n_i \geq 5$), prova-se que, sob H_0:

$$H \sim \chi^2_{k-1}$$

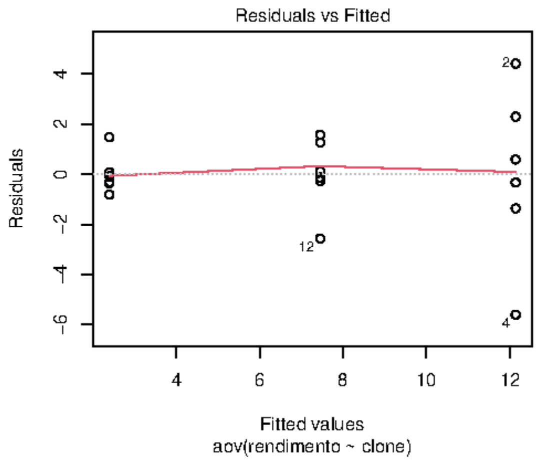
Região crítica: unilateral direita

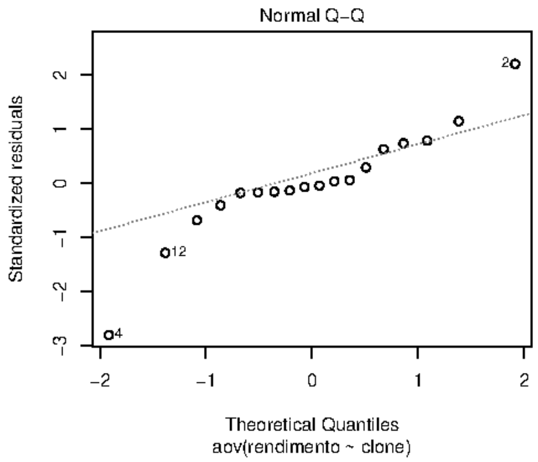
Para um nível de significância α, rejeita-se H_0 se $H_{calc} > \chi^2_{\alpha(k-1)}$

No R,

```
> kruskal.test (…)
```

Exercício 1: Num estudo sobre clones da variedade de videira Grenache, estudaram-se 3 clones relativamente ao rendimento (kg/planta) num ensaio com um delineamento totalmente casualizado, com 6 repetições de cada clone. Após o ajustamento de uma análise de variância clássica a um factor de efeitos fixos, a análise dos resíduos evidenciou a violação dos pressupostos da homogeneidade de variâncias e normalidade.





Um analista sugere a realização de um teste de Kruskal-Wallis. Os dados da experiência encontram-se seguidamente descritos.

clone	rendimento	**ordem**
GR1	12.725	**16**
GR1	16.550	**18**
GR1	14.432	**17**
GR1	6.538	**8**
GR1	10.775	**14**
GR1	11.808	**15**
GR2	9.025	**13**
GR2	7.575	**11**
GR2	7.317	**10**
GR2	7.188	**9**
GR2	8.717	**12**
GR2	4.888	**7**
GR3	2.317	**4**
GR3	2.088	**3**
GR3	3.875	**6**
GR3	2.045	**2**
GR3	2.475	**5**
GR3	1.588	**1**

a) Enuncie as hipóteses a testar no contexto deste estudo.

b) Calcule a estatística do teste de Kruskal-Wallis para a amostra observada.

c) Para um nível de significância (α) de 0,05, qual a conclusão que tira sobre o estudo efectuado?

Exercício 2: Num estudo sobre variedades tradicionais de alho, estudaram-se 4 variedades relativamente a características do bolbo. Concretamente, fez-se um ensaio com um delineamento totalmente casualizado com 5 repetições de cada variedade e em cada parcela avaliou-se o número de dentes presentes em 10 bolbos. Os valores obtidos em cada parcela, bem como o resultado obtido, no R, com o comando **kruskal.test(dentes~variedade, data=alhobd),** são seguidamente apresentados:

variedade	Nº dentes em 10 bolbos	ordem
V1	59	1
V1	65	5
V1	72	8
V1	81	11
V1	89	16
V2	61	2
V2	62	3
V2	73	9
V2	82	12
V2	90	17
V3	75	10
V3	88	15
V3	94	18
V3	96	19
V3	98	20
V4	63	4
V4	66	6
V4	67	7
V4	83	13
V4	87	14

```
> kruskal.test(dentes~variedade, data=alhobd)

        Kruskal-Wallis rank sum test

data:  dentes by variedade
Kruskal-Wallis chi-squared = 6.6571, df = 3, p-value = 0.08367
```

a) Enuncie as hipóteses a testar no contexto deste estudo.

b) Verifique que o valor da estatística do teste de Kruskal-Wallis para a amostra observada é 6.6571.

c) O que conclui sobre o estudo efectuado?

Nota:

Teste de Friedman

[Friedman , M. (1937). The use of ranks to avoid the assumption of normality in the analysis of variance. *J. Amer. Statist. Assoc.* 32 : 675 – 701.]

Variante não paramétrica de uma ANOVA a 2 fatores factorial , sem repetições nas células (sem interação)

No R,

```
> friedman.test (…)
```